

Ficha — Distribuição de amostragem e o Teorema Limite Central

11.º Ano

MACS

4 páginas

12 exercícios

Objetivo: Compreender a **distribuição de amostragem** de uma estatística (e a da média), relacionar a sua **variabilidade com a dimensão** das amostras e aplicar o **Teorema Limite Central** — além de distinguir o desvio padrão amostral **s** (**n – 1**) de **s'** (**n**).

Revisão rápida

Distribuição de amostragem = distribuição dos valores de um **estimador** quando se consideram **todas as amostras possíveis** (da mesma dimensão), com um determinado esquema de amostragem. Não é o histograma da amostra — é a distribuição **do estimador entre amostras**.

Regra	Leitura
Depende da dimensão n	tanto menor variabilidade quanto maior a amostra
Centro	em torno do parâmetro ($\bar{x}$ oscila em torno de μ)
TLC	para n suficientemente grande , o modelo Normal aproxima a distribuição de amostragem da média — qualquer que seja a forma da população

Porque é «limite central»: em populações grandes ou infinitas, **enumerar** todas as amostras seria «uma tarefa a maior parte das vezes impossível» (documento) — o TLC **resolve sem contar**: com n grande, a distribuição de $\bar{x}$ tem forma de sino, **independentemente da origem** (poissão, exponencial, o que for) — e é isso que legitima usar $\bar{x}$ para estimar μ .

Dois desvios-padrão amostrais (para estimar σ):

$$s = \sqrt{(\sum(x_i - \bar{x})^2 / (n - 1))} \cdot s' = \sqrt{(\sum(x_i - \bar{x})^2 / n)}$$

Usa-se **s** (**n – 1**) porque fornece **melhores estimativas** de σ (corrige a subestimação: a média $\bar{x}$ já «roubou» um grau de liberdade) — **s'** tende a ser menor que o verdadeiro σ .

Estratégia: reconhecer se é «distribuição de amostragem» (do estimador, entre amostras) → ver n (variabilidade?) → aplicar TLC (n grande → Normal) → usar **s** para descrever a amostra.

Exercícios

1. Completar: a distribuição de amostragem é a distribuição do _____ em todas as _____ possíveis de mesma _____ · a variabilidade é menor para n _____ · o TLC aproxima por uma _____ para n suficientemente grande.

2. Reconhecer: a distribuição de **p̂** da ficha 10 (3 mulheres + 2 homens, n = 2: 0,3 / 0,6 / 0,1) — (a) é uma distribuição de amostragem? Porquê (estimador, todas as amostras)? (b) O que seria um erro: chamar «distribuição de amostragem» ao gráfico de idades de uma só turma?

3. Construir com contas: população {2, 4, 6} ($\mu = 4$), amostras de dimensão 2 com reposição (9 amostras). (a) Enumera as 9 e os $\bar{x}$. (b) Distribuição de $\bar{x}$ (1/9, 2/9, 3/9, 2/9, 1/9). (c) Calcula a média da distribuição e compara com μ .

4. Dimensão e variabilidade: população $\{2, 4, 6\}$ com σ da população $\approx 1,63$. (a) Com $n = 1$, que distribuição tem $\bar{x}$? (b) Com $n = 4$ (reposição), escreve as amostras «todas iguais» (2222, 6666...) e observa: que valores extremos ficaram impossíveis? (c) Conclusão sobre σ de $\bar{x}$ com n a crescer.

5. O enunciado do TLC: completa — «para amostras aleatórias de dimensão _____, o modelo _____ aproxima a distribuição de amostragem da média, _____ que seja a forma da _____ subjacente» — e explica o que isto autoriza (estimar μ com $\bar{x}$? desenhar sinos?).

6. Porque «limite central»: (a) se a população for em forma de U ou exponencial, o que acontece à distribuição de $\bar{x}$ com $n = 100$? (b) E com $n = 3$? (c) O que «centraliza» — a média de quantos efeitos? Explica a ideia: cada $\bar{x}$ é uma média de n valores, e as médias de muitos valores tendem para o sino.

7. s vs s' : para os dados $\{2, 4, 6\}$ ($\bar{x} = 4$, $\sum(x_i - \bar{x})^2 = 8$): (a) calcula $s' = \sqrt{8/3} \approx 1,63$ e $s = \sqrt{8/2} = 2$. (b) Qual se usa para estimar σ e porque (subestimação com n)? (c) Numa calculadora, ao veres «desvio padrão» de uma amostra, que das duas é (o padrão é s , com $n - 1$)?

8. Do impossível ao teorema: população de 1 000 pessoas, amostras de $n = 30$. (a) Quantas amostras há (ordem de grandeza: combinações gigantes)? (b) Porque é que, apesar disso, podemos falar com segurança sobre a distribuição de $\bar{x}$? (c) O que o TLC dispensa (enumerar? gerar?).

9. Erros: corrige (a) «o TLC diz que a população é normal» · (b) «n = 3 amostras já servem para aplicar o TLC» · (c) «s' é melhor que s» · (d) «distribuição de amostragem = histograma dos dados da amostra».

10. $\sigma/\sqrt{n}$ em ação: a dispersão típica de $\bar{x}$ é $\sigma/\sqrt{n}$. Com $\sigma = 10$: (a) calcula-a para $n = 4$, $n = 25$ e $n = 100$. (b) O que acontece quando n quadruplica (de 4 a 16)? (c) Interpreta: porque é que «mais 75 questionários» quase não mexem na margem (a raiz engole o ganho)?

11. Tecnologia (AT 16): descreve como ver o TLC numa folha de cálculo — sortear 30 amostras de uma população em forma de exponencial, calcular $\bar{x}$ de cada, histogramar os $\bar{x}$ e observar a forma; repetir com $n = 3$. Que mudança de forma se regista e o que ela ilustra?

12. Síntese: «O TLC é o que torna a inferência possível.» Explica (4–5 linhas) o problema (enumerar amostras é impossível), a solução (com n grande, $\bar{x}$ segue Normal sempre, qualquer que seja a população) e a consequência prática — a normal da ficha 9 «volta» e permite construir intervalos de confiança para μ (ficha 12).

Soluções — Distribuição de amostragem e TLC

11.º Ano

MACS

Resolução completa

1. estimador · amostras · dimensão · maior · Normal.

2. (a) **Sim**: p^* é um **estimador** e a tabela 0,3/0,6/0,1 foi obtida com **todas as 10 amostras** de dimensão 2 — é a distribuição do estimador entre amostras (e a média dela, 0,4, acerta p). (b) Seria erro porque o gráfico de idades de uma **única** amostra descreve os **dados**, não o **comportamento de um estimador entre amostras** — falta o «e se repetíssemos a amostra?».

3. (a) 9 amostras: (2,2)→2 · (2,4)→3 · (2,6)→4 · (4,2)→3 · (4,4)→4 · (4,6)→5 · (6,2)→4 · (6,4)→5 · (6,6)→6. (b) $\bar{x} = 2 \cdot (1/9) + 3 \cdot (2/9) + 4 \cdot (3/9) + 5 \cdot (2/9) + 6 \cdot (1/9) \rightarrow$ soma 9/9 ✓. (c) Média = $(2 + 2 \cdot 3 + 3 \cdot 4 + 2 \cdot 5 + 6)/9 = (2+6+12+10+6)/9 = 36/9 = 4 = \mu$ ✓ (não-enviesada) — e repare que a distribuição é **simétrica em torno de μ** , mesmo com população discreta.

4. (a) Com $n = 1$, $\bar{x}$ = o próprio valor sorteado → {2, 4, 6} com probabilidades 1/3 cada (a distribuição da população). (b) Com $n = 4$, os extremos «só 2» e «só 6» puseram-se quase impossíveis ($2222 = (1/3)^4 \approx 1,2\%$); os $\bar{x}$ concentram-se perto de 4 — misturar 4 valores **lava** os extremos. (c) Conclusão: a dispersão de $\bar{x}$ **diminui com n** (σ de $\bar{x} = \sigma/\sqrt{n} \rightarrow$ cai) — mais amostra, estimativa mais estável (ex. 10).

5. «dimensão **suficientemente grande**» · **Normal** · «qualquer que seja» · **população**. Autoriza: (1) modelar a distribuição de $\bar{x}$ como **$N(\mu, \sigma/\sqrt{n})$** (forma conhecida) e (2) usar a normal (ficha 9) para traduzir $\bar{x}$ em **afirmações sobre μ** — é a base dos intervalos de confiança (ficha 12).

6. (a) Com $n = 100$, a distribuição de $\bar{x}$ fica **aproximadamente normal** (sino), centrada na média da população — a forma «U» desaparece. (b) Com $n = 3$, ainda se vê **resquício** da forma original (sino «embutido» na U — não basta). (c) Cada $\bar{x}$ é a **média de n valores** — ao combinar muitos valores, os extremos se anulam (uns puxam para cima, outros para baixo) e o resultado concentra-se em torno da média: é a «**média de médias**» que se centraliza (daí «limite central»).

7. (a) $s' = \sqrt{8/3} = \sqrt{2,667} \approx 1,63$ · $s = \sqrt{8/2} = \sqrt{4} = 2$. (b) Usa-se **s** (com $n - 1$): como $\bar{x}$ é estimada **dos próprios dados**, os desvios ($x_i - \bar{x}$) ficam artificialmente menores (roubaram-se graus de liberdade) — dividir por $n - 1$ corrige e dá valor **sem viés** de σ ; s' tende a subestimar. (c) A calculadora mostra **s** (a de amostra, $n - 1$) — a de população (σ) é raramente dada.

8. (a) $C(1000, 30) \approx 10^{87}$ — número inconcebível (impossível enumerar). (b) Porque o **TLC** diz a forma da distribuição **sem enumerar**: com $n = 30$ (n grande, regra prática), $\bar{x}$ segue aproximadamente **Normal** — a segurança vem do teorema, não da contagem. (c) O TLC dispensa **enumerar/gerar** as amostras — dispensa o «impossível» do ex. 9 da ficha 10.

9. (a) **Falso** — o TLC afirma que a distribuição **de $\bar{x}$** (o estimador) tende para a Normal; a população pode continuar como está. (b) **Falso** — exige **n suficientemente grande** (regra prática: dezenas — $n = 3$ mantém a forma original). (c) **Falso** — para estimar σ usa-se **s ($n - 1$)** (s' subestima). (d) **Falso** — o histograma da amostra é dos **dados**; a distribuição de amostragem é das **estatísticas de cada amostra** ($\bar{x}$ de $\bar{x}$ de $\bar{x}$...).

10. (a) $\sigma/\sqrt{n}$: $n = 4 \rightarrow 10/2 = 5$ · $n = 25 \rightarrow 10/5 = 2$ · $n = 100 \rightarrow 10/10 = 1$. (b) De 4 a 16, a dispersão passa de 5 para 2,5 (**divide-se por 2** — quadruplicar n divide por $\sqrt{4}$). (c) A raiz **engola** o ganho: para reduzir pela metade é preciso **quadruplicar** n (de 25 a 100 para ir de 2 para 1) — é por isso que os inquéritos «maiores» custam muito e melhoram pouco: o custo cresce linearmente, o ganho, pela raiz.

11. (1) Ter a população (ex.: valores em exponencial) e sortear **30 amostras** de $n = 30$ (função aleatória) → $\bar{x}$ de cada. (2) **Histogramar os 30 $\bar{x}$** . (3) Repetir com **$n = 3$** e comparar: com $n = 30$ os $\bar{x}$ formam um **sino** (forma original invisível); com $n = 3$ o histograma **assemelha-se à população** (assimétrico). Ilustra o TLC: a normalidade «acende-se» com n grande — e a dispersão também cai ($\sigma/\sqrt{n}$).

12. O problema: a distribuição de amostragem exige **todas as amostras** — em populações grandes é «impossível» ($C(1000,30) \approx 10^{87}$). A solução: o **TLC** — com n suficientemente grande, $\bar{x}$ segue **Normal sempre**, venha de que população vier (poisson, exponencial, U...), porque cada $\bar{x}$ é uma média de n efeitos que se anulam. A consequência: a **normal da ficha 9 reaparece** onde não existia — e com **$N(\mu, \sigma/\sqrt{n})$** dá-se o salto decisivo: transformar $\bar{x}$ em **intervalo para μ** com confiança declarada (ficha 12) — do teorema ao juízo.

Nota: quatro blocos — **definir distribuição de amostragem · dimensão e variabilidade · TLC (enunciado e ideia) · s vs s'** . Liga: ficha 9 (normal) · ficha 10 (amostragem) · ficha 12 (IC) · AT 16 (simulação do TLC) · resumo 3.